412TW-PA-12811

Confidence Intervals for Binary Responses- R50 & the Logistic Model

4
1
2
T
W

ARNON M. HURWITZ

AIR FORCE TEST CENTER
EDWARDS AFB, CA

OCTOBER, 2012

AIR FORCE TEST CENTER
EDWARDS AIR FORCE BASE, CALIFORNIA
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
October 2013	Presentation	1 October 2013 – 30-October 2013

4. TITLE AND SUBTITLE

Title: Confidence Intervals for Binary Responses
Subtitle: R50 & the Logistic Model

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Arnon M. Hurwitz (US Air Force)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) AND ADDRESS(ES)

Air Force Flight Test Center
412 Test Wing
Edwards AFB CA 93524

8. PERFORMING ORGANIZATION REPORT NUMBER

412TW-PA-12811

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AIR FORCE TEST CENTER
EDWARDS AFB, CA

10. SPONSOR/MONITOR'S ACRONYM(S)

N/A

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release A: distribution is unlimited.

13. SUPPLEMENTARY NOTES

CA: Air Force Test Center Edwards AFB CA CC: 012100

14. ABSTRACT

Logistic regression is a non-linear method for modeling a binary response variable. For example, y = {success, failure} for blip-scan radar detections. Such responses cannot be modeled using regular linear regression. In our work, many applications of logistic regression present themselves. In the present discussion, models allowing independent slopes and independent intercepts are considered for comparing multiple groups of measures. The question that we consider here is the construction of a confidence interval about the difference in the radar 'Range 50' (R50) values for two logistic curves with each value (viz. R1, R0) arising from the separate curve. R50 represents the range at which radar achieves 50% detection probability. This problem is the same as the problem of prediction of the LD50 ('lethal dose/effective dose 50 %') value in medical science. We approach the problem analytically using parametric methods. A feature is the use of 'inverse prediction' or calibration methods. Our results are based on the large-sample properties of Maximum Likelihood estimation, and improve on results based on the least-squares model. The application is also given for general Rp/Lp —that is, range/dose values not equal to R50. Results for large and small samples are checked against a 'truth source' generated using a Bootstrap program

15. SUBJECT TERMS

Logistic, radar, R50, LD50, calibration, inverse prediction, confidence interval

16. SECURITY CLASSIFICATION OF: **Unclassified**			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			412 TENG/EN (Tech Pubs)
Unclassified	Unclassified	Unclassified	None	14	19b. TELEPHONE NUMBER *(include area code)* 661-277-8615

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

War-Winning Capabilities … On Time, On Cost

Confidence Intervals for Binary Responses- R50 & the Logistic Model

ACAS, October 2012. Monterey, CA

Dr. Arnon Hurwitz - Edwards AFB, CA 93524

Contact: arnon.hurwitz@edwards.af.mil

412th Test Wing

Confidence Intervals for Binary Responses
ACAS 2012, Monterey, CA

U.S. AIR FORCE

Statistical Methods Group, Edwards AFB

AIR FORCE MATERIEL COMMAND

Overview

- Blip-scan radar output returns are: detect/no-detect, or {0, 1}
- Probability of detection **π** increases as range-to-target decreases
- A common metric is R50 – the range at which **π** = 50%
- A common question is: given two flights, what is a confidence interval (C.I.) for the *difference of the two R50's ?*
- Such R(π) differences are non-linear functions of the parameters of the estimation procedure; its own distribution is hard to derive
- A solution to find a C.I. for a difference is to use a Bootstrap procedure = a non-parametric simulation approach
- Bootstrapping works, but it has to be custom-generated for each different problem at hand. It's sometimes preferable to have a parametric method. We develop such a method here based on the Max. Likelihood Covariance (inverse of the Fisher Information).

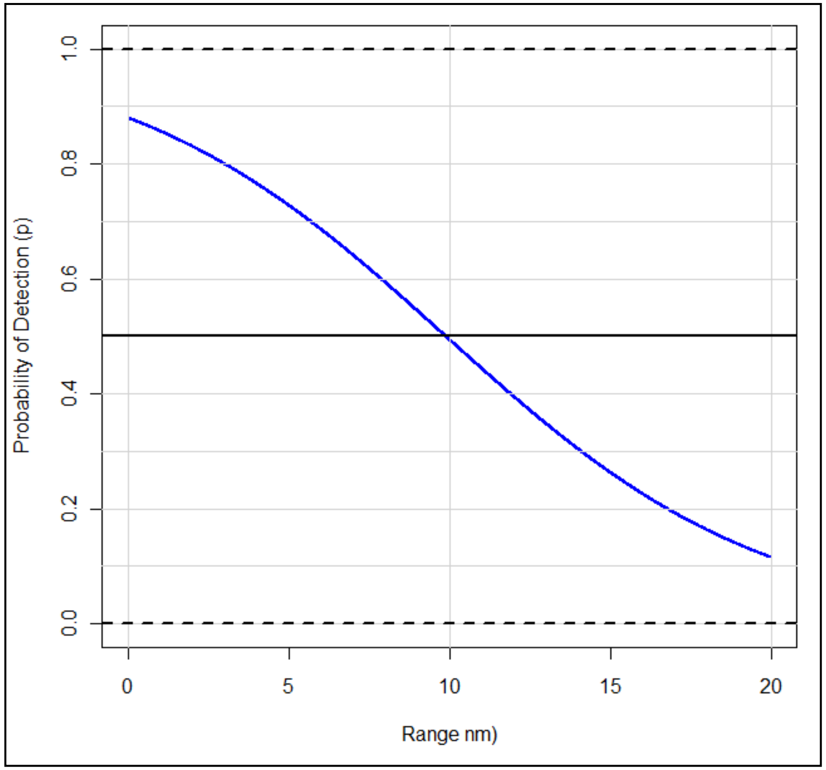
Logistic curve fit to Binary data

- **One has the relation:**
  - Output = function(Range)
- **But output is binary {0, 1}, and we'd rather wish to find something like:**
  - π = function(Range)
- **Transform the problem:**

  $y(R) = \log [\pi/(1 - \pi)] = \alpha + \beta R$

- **Now we have a linear relation of a kind, with**

  $\pi = \exp(y) / [1 + \exp(y)]$

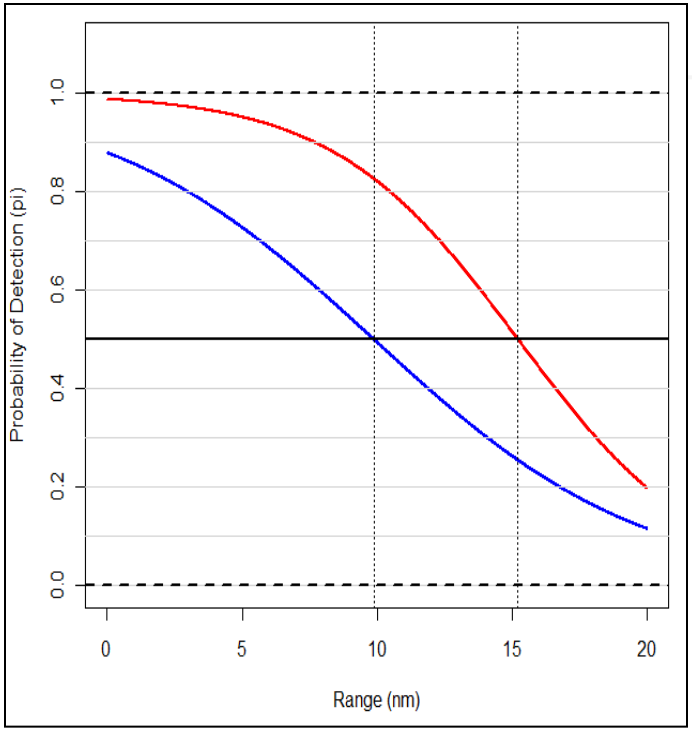


A Logistic curve

Probability is on the vertical axis

Comparing two logistic curves

- **log** [π/(1 - π)] == logit(π) **is called the '**logit**'. Log ==ln**
- **The graph shows two such curves (two flights)**
- **At π = 0.5 the blue curve shows R50 at about 10 nm, the red curve at about 15 nm**
- **The difference is about 5: we want a 95% confidence interval around this difference**



Two logistic curves

Here, P(detect) decreases with increasing R

Estimation

- **Let one curve be estimated with non-linear regression techniques (*generalized linear modeling*) to give the equation**

$$logit(\pi) = \alpha_0 + \alpha_1 R$$

and let the other curve be estimated as

$$logit(\pi) = \beta_0 + \beta_1 R$$

- **At π = 0.5, logit(π) = log(0.5/0.5) = log(1) = 0.**

So for the first curve R0 = $-\widehat{\alpha}_0/\widehat{\alpha}_1$**, and R1=** $-\widehat{\beta}_0/\widehat{\beta}_1$ **for the second.**
Their estimated difference is therefore R0 – R1 = $-(\hat{\alpha}_0/\hat{\alpha}_1 + \hat{\beta}_0/\hat{\beta}_1)$

- Generalized linear modeling uses **maximum likelihood estimation (MLE)** techniques to estimate the coefficients of the models, and also gives us the **Covariance Matrix of the α and β parameters**

- **Call this covariance matrix V. It is a 4x4 symmetric matrix.**

Confidence Interval

- It can be shown (by MLE large-sample theory) that

$$(\widehat{R1} - \widehat{R0}) \dot{\sim} \textbf{Normal(R1 – R0, hVh')}$$

Where V is the covariance matrix, and where

$$\mathbf{h} = \left(\frac{-1}{\alpha_1}, \frac{\alpha_0}{{\alpha_1}^2}, \frac{1}{\beta_1}, \frac{-\beta_0}{{\beta_1}^2}\right)$$

This gives us the (95%) confidence interval that we desire as:

$$(\widehat{R1} - \widehat{R0}) - 1.96 \times \hat{\mathbf{h}}\hat{\mathbf{V}}\hat{\mathbf{h}}' < \text{R1} - \text{R0} < (\widehat{R1} - \widehat{R0}) + 1.96 \times \hat{\mathbf{h}}\hat{\mathbf{V}}\hat{\mathbf{h}}$$

C.I. for the General Case

- So far, we've developed a CI for R50; that is, where π = 0.5
- We can get a CI for any value of π in (0, 1) by replacing the 'h' we used in the above slide with

$$\mathbf{h} = \left(\frac{-1}{\alpha_1}, \frac{-(y_c - \alpha_0)}{{\alpha_1}^2}, \frac{1}{\beta_1}, \frac{(y_c - \beta_0)}{{\beta_1}^2}\right),$$

where y_c is the estimate of the logit(π) at the new value of π.

- The above theory depends on the assumption that the two flights gave independently- estimated curves, and the curves do not cross over each other.

Bootstrap Test

We ran bootstrap simulations against our analytic technique for π = p = 0.2, 0.5, and 0.8 to see how we compared, and also compared our F.I. method against Schwenke & Milliken's

At probability	Method 1	Method 2	Method 3	Method 4
P	Analytical 95% C.I. based on Fisher Information	Bootstrap 95% C.I.	Bias Corrected Bootstrap 95% C.I.	Normal approx. method of Schwenke & Milliken
0.2	[2.7, 3.7]	[2.5, 3.8]	[2.56, 3.9]	[2.4, 4.0]
0.5	[4.8, 5.9]	[4.8, 5.9]	[4.9, 5.9]	[4.5, 6.2]
0.8	[6.9, 8.0]	[6.7, 8.3]	[6.7, 8.3]	[6.7, 8.3]

Our method produces intervals very close to the Bootstrap ☺

Note: These are all large-sample results

Small-sample Test

We looked at a smaller sample size (n=200), and compared our 'Analytical' method against Schwenke & Milliken's. (The data is randomly sampled from the original bootstrap data set).

p =0.5 n=200	Method 1	Method 2	Method 3
Run	Analytical 95% C.I. based on Fisher Information	Schwenke & Milliken's method 95% C.I.	Bias Corrected Bootstrap 95% C.I.
#1	[0.4, 5.5]	[-1.8, 7.8]	[4.9, 5.9]
#2	[3.8, 8.1]	[1.2, 10.8]	[4.9, 5.9]
#3	[2.1, 6.0]	[0.5, 7.6]	[4.9, 5.9]

Our method produces narrower intervals than S&M

NOTE: This conclusion is based on just 3 runs; more extensive tests are planned

Summary- R50 & the Logistic Model

- We looked at a CI on the difference between R50 points for two independent flights
- We extended the results to the difference between two Rp points, where $0 < p < 100$
- Our 'analytic' method is based on the covariance matrix generated from the MLE procedure of generalized regression
- We compared CI's of our method to the 'true' CI generated by a large bootstrap sample (n = 4000 scans/flight), and also to an alternate method by Schwenke & Milliken (1991)
- We further looked at the comparative results for a 'small' sample (n = 200 scans/flight).

REFERENCES

- [1] Schwenke, J.R. and Milliken, G.A. (1991). ***On the Calibration Problem Extended to Nonlinear Models.*** **Biometrics 47, 563-574**

- **[2] McCullagh P. and Nelder, J. A. (1989).** ***Generalized Linear Models.*** **Chapman & Hall**

- **[3] Kendall, M.G. and Stuart, A.S. (1961).** ***The Advanced Theory of Statistics. Vol.1*****. 2nd edition. Griffin**

- **[4] Efron, B. and Tibshirani, R.J. (1993). An Introduction to the Bootstrap. Chapman & Hall.**

- **[5] Hurwitz, A. (2012).** ***Confidence Intervals for Binary Responses – R50 & the Logistic Model*****. (AFTC: To be released)**